

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 19:44:29 ; Search time 30.3429 Seconds
(without alignments)
3178.015 Million cell updates/sec

Title: US-09-497-967-7
Perfect score: 2540
Sequence: 1 MKNILVILISLFIQIKS.....QCDFANFLSLLLSIYLL 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhmc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirls:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	2540	100.0	468	5 Q9BMH3	Q9bmh3 ichthyophth
2	921	36.3	442	5 Q9XZG2	Q9xzg2 ichthyophth
3	878	34.6	460	5 Q962N5	Q962n5 ichthyophth
4	775.5	30.5	395	5 Q27208	Q27208 ichthyophth
5	345	13.6	371	5 Q9GPP0	Q9gpp0 tetrahymena
6	289.5	11.4	316	5 Q9GPP3	Q9gpp3 tetrahymena
7	286.5	11.2	315	5 Q9GPP4	Q9gpp4 tetrahymena
8	283.5	11.2	305	5 Q9GPP2	Q9gpp2 tetrahymena
9	236	9.3	594	5 Q24970	Q24970 giardia lam
10	232	9.1	645	5 Q97448	Q97448 giardia lam
11	228	9.0	1274	5 Q24977	Q24977 giardia lam
12	227.5	9.0	667	5 Q9XTK3	Q9xtk3 giardia lam
13	226	8.9	719	5 Q9U019	Q9u019 giardia lam
14	221.5	8.7	719	5 Q9U021	Q9u021 giardia lam
15	219	8.6	1274	5 Q9NGL3	Q9ngl3 giardia lam
16	218.5	8.6	548	5 Q9GQ45	Q9gq45 giardia lam

17	218.5	8.6	1372	5 P91526	P91526 caenorhabdi
18	218	8.6	1154	5 Q9GQ46	Q9gq46 giardia lam
19	216.5	8.5	436	5 Q27197	Q27197 tetrahymena
20	216.5	8.5	597	5 Q07317	Q07317 giardia int
21	210.5	8.3	560	5 Q2U013	Q2u013 giardia lam
22	210.5	8.3	769	5 Q24971	Q24971 giardia lam
23	209	8.2	704	5 Q9U048	Q9u048 giardia lam
24	206	8.1	3396	5 Q9VM55	Q9vm55 drosophila
25	205	8.1	709	5 Q9XTJ7	Q9xtj7 giardia lam
26	204	8.0	667	5 Q95WU1	Q95wu1 giardia lam
27	202	8.0	1551	5 Q9NGV4	Q9ngv4 drosophila
28	200	7.9	1704	5 Q94446	Q94446 chironomus
29	199	7.8	709	5 Q97444	Q97444 giardia lam
30	198	7.8	394	5 Q9GQ47	Q9gq47 giardia lam
31	197	7.8	557	5 Q24992	Q24992 giardia lam
32	196	7.7	503	5 Q9U018	Q9u018 giardia lam
33	193	7.6	556	5 Q9NGZ3	Q9ngz3 giardia lam
34	191.5	7.5	704	3 Q74567	Q74567 trichoderma
35	191.5	7.5	1297	5 Q26632	Q26632 strongyloce
36	190.5	7.5	421	5 Q95V69	Q95v69 tetrahymena
37	190.5	7.5	2972	5 P90891	P90891 caenorhabdi
38	189	7.4	1101	5 Q964D2	Q964d2 entamoeba h
39	189	7.4	1622	5 Q06550	Q06550 cryptospori
40	189	7.4	5374	11 Q95ND0	Q95nd0 mus musculu
41	188.5	7.4	350	5 Q94589	Q94589 leishmania b
42	188.5	7.4	2759	5 Q45614	Q45614 caenorhabdi
43	188.5	7.4	3102	5 Q9TZR4	Q9tzr4 caenorhabdi
44	188	7.4	739	5 Q9GS24	Q9gs24 giardia lam
45	188	7.4	1074	5 Q964D1	Q964d1 entamoeba h

ALIGNMENTS

RESULT 1

Q9BMH3 ID Q9BMH3 PRELIMINARY; PRT; 468 AA.
AC Q9BMH3;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DE 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE Immobilization antigen isoform.
GN IAG52A.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G5;
RA Lin Y., Lin T.-L., Clark T.G.;
RT "Variation in primary sequence and tandem repeat copy number among i-
RT antigen genes of Ichthyophthirius multifiliis.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF324424; AAK01661.1; -
SQ SEQUENCE 468 AA; 48281 MW; BEAGDA42833A7726 CRC64;

Query Match 100.0%; Score 2540; DB 5; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.1e-177;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNILVILISLFIQIKSANGCPVGTETNTAGQVDDLTGTPANCVCNCKNFYNNAAAFV 60
Db 1 MKNILVILISLFIQIKSANGCPVGTETNTAGQVDDLTGTPANCVCNCKNFYNNAAAFV 60
|||||

QY 61 PGASTCTPCQKKDAGAQPNPATANLVTCNVKCPAGTAAGATDYAAIITECVNCRI 120
Db 61 PGASTCTPCQKKDAGAQPNPATANLVTCNVKCPAGTAAGATDYAAIITECVNCRI 120
|||||

QY 121 NFYNENAPNFAGASTCTCTCPVNRVGGALTAGNAATIVACQNCPTGALDDGVTTDYY 180
Db 121 NFYNENAPNFAGASTCTCTCPVNRVGGALTAGNAATIVACQNCPTGALDDGVTTDYY 180
|||||

QY	181	RSFTECVKRLNFYYNGNGNTFFNPGKSGQCTPCPAIKPANTVAQATLGNDAITTAQC	NWA	240
Db	181	RSFTECVKRLNFYYNGNGNTFFNPGKSGQCTPCPAIKPANTVAQATLGNDAITTAQC	NWA	240
QY	241	CPDGTISAAGVNNWVAQNTCTNCAPNFYNNAPNPNPGNSTCLPCPANKDYGA	EATAGG	300
Db	241	CPDGTISAAGVNNWVAQNTCTNCAPNFYNNAPNPNPGNSTCLPCPANKDYGA	EATAGG	300
QY	301	AATLAKQCNIAACPDGTAIASGATNYYVILQTECINCAANFYFDGNNFQAGSSRCKAC	PANK	360
Db	301	AATLAKQCNIAACPDGTAIASGATNYYVILQTECINCAANFYFDGNNFQAGSSRCKAC	PANK	360
QY	361	VOGAVATAGGTATLIAQCALECPAGTVILDTGTTSTYKQAASCEVCVCAANFYTTK	OTDWA	420
Db	361	VOGAVATAGGTATLIAQCALECPAGTVILDTGTTSTYKQAASCEVCVCAANFYTTK	OTDWA	420
QY	421	GIDTCTSCNKKLTSGAEANLPESAKKNIOCFANFLSISLLLSIYYLL	468	
Db	421	GIDTCTSCNKKLTSGAEANLPESAKKNIOCFANFLSISLLLSIYYLL	468	
RESULT 2				
Q9XZG2	ID	Q9XZG2	PRELIMINARY; PRT; 442 AA.	
AC	Q9XZG2;			
DT	01-NOV-1999	(T-EMBLrel. 12, Created)		
DT	01-NOV-1999	(T-EMBLrel. 12, Last sequence update)		
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)		
DE		Immobilization antigen precursor.		
GN		IAG48.		
OS		Ichthyophthirius multifiliis.		
OC		Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;		
OX		Ophryoglenina; Ichthyophthirius.		
OC		NCBI_Taxid=5932;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN=G1.		
RX		MEDLINE=99196987; PubMed=10095108;		
RA		Clark T.G., Lin T.B., Jackwood D.A., Sherrill J., Lin Y.,		
RA		Dickerson H.W.;		
RT		"The gene for an abundant parasite coat protein predicts tandemly		
RL		repetitive metal binding domains.";		
RE		Gene 229:91-100(1999).		
RN		[2]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN=G1;		
RA		Gaertig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;		
RT		"Surface display of a parasite antigen in the ciliate Tetrahymena		
RL		thermophila.";		
RE		Nat. Biotechnol. 0:0-0(1999).		
DR		EMBL; AF140273; AAD31283.1; -.		
DR		PRINTS; PR01574; TUBBYPROTEIN.		
KW		Signal.		
FT		SIGNAL	1 20	POTENTIAL.
FT		CHAIN	21 442	IMMOBILIZATION ANTIGEN.
SQ		SEQUENCE	442 AA; 45025 MW; 52658F3F65D27AFA	CRC64;
Query Match			36.3%; Score 921; DB 5; Length 442;	
Best Local Similarity			41.8%; Pred. No. 2.3e-59;	
Matches			214; Conservative 45; Mismatches 139; Indels 114; Gaps	
QY	1	MKNNILIILISIFINOIKSANGCVGTETNTAGQVD----	DLGT	PANCYNCKNFYNNNA 56
Db	1	MKNNILIILISIFINELRAVPCPDGTQTQ-AGLTDVGAADLGT----	CVNC	RPNFYINGG 56
QY	57	AAFPVGASTCTPCQKKDACAQNPNPATANLVTCQNVKPCAGTAIAGGATDYAAIITECV	116	
Db	57	AA-----QGEANGNQPPAAN-----	-----	71
QY	117	NCRFYNENAPNPNAGASCTCTACPVNVRGALTAGNAATIVTAQCNVACPTGTALDDGVT	176	
Db	72	-----NAARGICVPCQINRVGSVFNAGDLATLTAQCSTQCTGTALDDGVT	117	

```

Db 268 -----PGNSKCVACESKKT-NSQSRSGLEANLAAQCCTECPAGTL 307
QY 318 IASGAT-NYVILQTECLNCAANFYEDGNFQAGSSRCACAPANKVOGAVATAGGTATLIA 376
Db 308 VTDGVTPTVTSQVCNCKAGFY-QNSNFEAGKQCKNCAVSKT-GSASVPGNSATSAT 365
QY 377 QCALECPAGTTLTGTSTYKQAASECVKCAANFYTTKOTDHWAGIDTCTSCNKKLTSGA 436
Db 366 QCONDCPAGTVDDGTSTNFVALASECTKQCNFYASTSGTGAAGTDTCTECSKKLTSGA 425
QY 437 EANLPESAKKNTQC---DFANFLSLLLIISYLL 468
Db 426 TAKVVAETQKAQCASSSTFAKELSLIFISPYLL 460

RESULT 4
Q27208
ID Q27208 PRELIMINARY; PRT; 395 AA.
AC Q27208;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE Immobolization antigen precursor (Fragment).
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=92335298; PubMed=1631132;
RA Clark T.G., McGraw R.A., Dickerson H.W.;
RT "Developmental expression of surface antigen genes in the parasitic
RT ciliate Ichthyophthirius multifiliis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=93020590; PubMed=1383510;
RA Lin T.L., Dickerson H.W.;
RT "Purification and partial characterization of immobilization antigens
RT from Ichthyophthirius multifiliis.";
RL J. Protozool. 39:457-463(1992).
[3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; M92907; AAC36158.1; -.
DR PRINTS; PR01574; TUBBYPROTEIN.
KW Signal;
FT NON_TER 1 1
FT SIGNAL <1 1 POTENTIAL.
FT CHAIN 2 395 IMMOBILIZATION ANTIGEN.
SQ SEQUENCE 395 AA; 39567 MW; 68DA2C790E4FD393 CRC64;

Query Match 30.5%; Score 775.5; DB 5; Length 395;
Best Local Similarity 44.2%; Pred. No. 7.9e+49;
Matches 175; Conservative 35; Mismatches 129; Indels 57; Gaps 16;

QY 93 VKCPAGTATAGGATDY-AAITECVNCRINFY-----NENAP--NFNAGASTCTA 139
Db 2 VPCPDGTQAGLTDVGAADLTGTCVNCRPNFYNGRAQGEANGQPAANNAARGICVP 61
QY 140 CPVNRVGGALTAGNATVAQCNVACPTGTDLDGVTYDVRSTFCYKCRNFYNGNN 199
Db 62 CQINRVGSVTNAGDLATLATOCTGCTGTDLDGVTDFDRSAAQCVKCKPNFYNGGS 121
QY 200 --GNTP-----KSCQTPCPAIPKPNVAQATLGNDAITITACQNVACPDG 244
Db 122 PGEAPGVQVFAAGAAAAGVAATVQCPVQLNK--NDSPATGAQAANLATQCSQCPTG 179
QY 245 TISAAGVNNVNAQNTNTE---CTNCAPNFYNN-----NAPN---FNPNG-----NSTC 283

```

```

Db 180 TVLDDGVT--LVFTSATLVCVCRPNFYNGSGPQSGAPGVQVFAAGAAAAGVAATVQSC 237
QY 284 LPCPANKDYGAETAGGAATLAKQNCIACPDGTGTAIASGAT-NYVILQTECLNCAANFYFD 342
Db 238 VPCQINKN-DSPAATGAQAANLATQCTGCTGTAIDQGVTLVFSNSSTQCSQCIANFYFN 296
QY 343 GNNFQAGSSRCACAPANKVOGAVATAGGTATLIAQCALECPAGTTLTGTSTYKQAASE 402
Db 297 G-NLEAGKSQCLKCPVSKTTTPAHA-PGNTATQATQCLTCTPAGTVDLDDGTSTNFVASATE 354
QY 403 CVKCAANFYTTKOTDHWAGIDTCTSCNKKLTSGAEA 438
Db 355 CTRCSAGFFASKTGTGTAGTDTCTECTCKLTSGATA 390

RESULT 5
Q9GPP0
ID Q9GPP0 PRELIMINARY; PRT; 371 AA.
AC Q9GPP0;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE Immobolization antigen LD (Fragment).
GN SERLD.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANF18211;
RX MEDLINE=20549003; PubMed=11095959;
RA Doerder F.P., Gerber C.A.;
RT "Molecular Characterization of the SerL Paralogs of Tetrahymena
RT thermophila.";
RL Biochem. Biophys. Res. Commun. 278:621-626(2000).
DR EMBL; AF312775; AAG38107.1; -.
FT NON_TER 1 1
SQ SEQUENCE 371 AA; 35175 MW; 5817EFFF2517DEAC CRC64;

Query Match 13.6%; Score 345; DB 5; Length 371;
Best Local Similarity 30.1%; Pred. No. 1.6e+17;
Matches 141; Conservative 41; Mismatches 175; Indels 112; Gaps 33;

QY 8 ILIISLFINQIKSAN-CPVGTETNTAGQVDDLTGPANVCNCKNFYNNAAAFVPGASTC 66
Db 6 LILISLAV--IATVNVAC---TDTNATA-----GAGGTCTF-CNAGYYGTSTDVTASGA--C 52
QY 67 TPCQKKDAGANPPATANLVTQCNVCKPAGTAIAGGATDYAAIITECVNCRINFYNNEN 126
Db 53 QKCPTGTNSVA---ATASGTLVTSCT---CNDTNAGLKADNSG-----COCKANFY--G 98
QY 127 APNENAGAST-CTACPVNRVGGALTAGNAATIVAAQCNVACPTGTALDDGVTYDVRSTFE 185
Db 99 TPNVAVAGATGCTTACP--TGTASPACTAAVTSCAN-----DTNASLRGDSN 143
QY 186 CVKCRNLNFYNGNNGTNPFNPGKSQCTPCPAIPKPNVAQATLGNDAITITACQNVACPDGT 245
Db 144 GCQCKANFYGTNP---AVAGGATGCTTACP-----TGSAAAAGSTAVTSCACN-----DT 189
QY 246 ISAAGVNNVNAQNTNTECNAPNFYNNNAPNPNPNCST-CLPCPANKDYGAETAGGAATL 304
Db 190 NSAL-----RADNSACI-CKANFY--GTPNAVAGGATGCTTACPT-----GSAAGASTAVT 237
QY 305 AKQCNIACPDGTATASGATNYVILQTECLNCAANFYFDGNFQAGSSRCACAPANKYQGA 364
Db 238 SCACN-----DTNSALKADN-----SACI-CKANFYCTPNVAVAGGATGCTTACPT---GT 282
QY 365 VATAGGTATLIAQCALECPAGTTLTGTSTYKQAASECVKCAANFYTTKOTDHWAGIDT 424
Db 283 TSTAG--TTVIGSCA--CP-----DTNASLNTATPVPQCNCANFYGTPTTTGASG--- 328

```


Query Match 11.2%; Score 283.5; DB 5; Length 305;
Best Local Similarity 30.1%; Pred. No. 4e-13;
Matches 118; Conservative 34; Mismatches 141; Indels 99; Gaps 29;

```
QY 8 ILIISLFINOIKSAN-CPVGTETNTAGVDLLCTPANCYNCKNFYNNAAAFVPGASTC 66
DB 6 LILISLAV--IATVNAC---TDTNATA-----GAGGTCTP-CNAG-YTGTSTDTVTPSGS-C 52
QY 67 TPCPQKKDAGQNPPTANLVTQCNVKKCPAGTAAGGATDYAAIIITECVNCRINFINEN 126
DB 53 TKCPTGNSVA---ATASGTLVSSCT--CNDTNASLKGDN-----SGC-QCKANFY--G 98
QY 127 APN-FNAGASTCTACPVNRRVGGALTAGNAATIQAOCNVACPTGALDDGGVTTDYVRSFTE 185
DB 99 TPNVAVGGATGCSACP---TGTTSPAGTAATVSCACN-----DTNASLKGDN 143
QY 186 CVKRLNFYNGNGNTPFPNGKSOCTPCPAIKPANVAQATLGNDAITTAOCNVACPDGT 245
DB 144 GCQCKANFYGTPN---AVAGGATGCTACTGSA-----AAGSTAVTSCANDTN 190
QY 246 ISAAGVNNVAQNTCTNCAPNFYNNAPNPNFNST-CLPCPANKDYGAETAGGAATL 304
DB 191 SSLK-----ADNSACY-KRANFY--GTPNAVAGGATGCTACTP-----GTTSTAG--TTV 235
QY 305 AKOCNIACPDGTATASGATNVVILQTECLNCAANFYFDGNNFOAGSSRCKACPANKVQGA 364
DB 236 IGSC--ACPDNNAALNATPPV-----COCKANFY--GTPTAGSAGCTACPS-----GQ 281
QY 365 VATAGGATLIAOCALECPAGTVLTDGTTSTY 396
DB 282 TAPA-GSATNVCKAA-----STSTSTY 301
```

RESULT 9

```
Q24970 ID Q24970 PRELIMINARY; PRT; 594 AA.
AC Q24970;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein.
GN VSPA6-S1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WBA6;
RX MEDLINE=96010579; PubMed=7581319;
RA Yang Y.M., Adam R.D.;
RT "Analysis of a repeat-containing family of Giardia lamblia variant-
RT specific surface protein genes: diversity through gene duplication and
RT divergence."
RL J. Eukaryot. Microbiol. 42:439-444 (1995).
DR EMBL; U17980; AAA82585.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00261; FU; 3.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ SEQUENCE 594 AA; 59575 MW; 9680818FB75F52AC CRC64;
```

Query Match 9.3%; Score 236; DB 5; Length 594;
Best Local Similarity 23.2%; Pred. No. 2.3e-09;
Matches 131; Conservative 47; Mismatches 207; Indels 180; Gaps 32;

```
QY 20 SANCVPVGTETNTAGVDLLCTPANCYNCKNF--YNNAAAFVPGASTCTPCPKKDAQ 78
DB 40 NGNTPYLKKTNPS---DPTGTCTVSAVDCQGSAGYYTDDS--VSDAKECKKC----- 85
```

```
QY 79 PNPP-----ATANLYTQC-----NVKCPAGTAIA-----GGA-----TDYAAII 112
DB 86 -NAPCTACAGTADKCTKCDANGAAPYLKKTNPSTGTCTCVSAVDCQGSAGIYTTDSVDA 144
QY 113 TCVNCRINFINENAPN-----FNAGASTCTCTACPVNRRVGGALTAGNAATIQAOCNVACP 166
DB 145 KECKKA---EGQKPTAGTQCFSCSDANCERCDDNDVCARCSTG-APPENGRKPAATP 199
QY 167 TGTALDDGVTTDYVRSFTE-CVKRLNFYNN-----GNNG-----NTPF----- 204
DB 200 GCHSSCDGCTENAMTNQADKCTGCKEGRYLKPESAAAGSGTCLTAECTSDTTHFTKEKA 259
QY 205 NFGKSOCTPCPAIKPA-----NVAQATLGNDAITTAOCNVACPDGTISAAGVNNVAQNT 259
DB 260 GDSKGMCLPCSDATHIAGCKKALKTLGSGAEASTVWCS-ECTDKWLTPSG-----NA 311
QY 260 ECTNCAPNFYNNAPNPNFNPNSTCLPCPANKD-----YGAETAGGAATIA 305
DB 312 CLDNCAPGYTPNDNLCTSDHTCAECNADRASCTACYPGYSLLYGS-CTAG---TCV 367
QY 306 KOCNTA-----CPDG-----TAI-ASGATNY-VILQTECL 333
DB 368 KECTGAFGANCADGGCTADVGGAKYCAQCKDGYAPIDGICTAVAAAGRTNVCTAADGTCT 427
QY 334 NCAANF-YFDGNF-----QAGSSRCKACPANKVOGAVATAGGTATLIAOAL 380
DB 428 KCAGEYTLMSGCYGVAKLPKGSVCTLASNGKICMAAN-----GOAPVQEKCP- 476
QY 381 ECPAG-----TVLTDGTTSTYKQAASECVKCAANFYTTKOTDWWAGIDTC----- 425
DB 477 ECSEGCACKCNDSNACTECLPGYYKGADKCFKCTASSGNNGNITGVANCVTCAPPAGSG 536
QY 426 -TSCNKKL--TSGAEANLPESAKKN 447
DB 537 SVTCYVKTGDTSGDDNDTGGSVNKS 561
```

RESULT 10

```
O97448 ID O97448 PRELIMINARY; PRT; 645 AA.
AC O97448;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine-rich protein.
GN CRP65.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96186899; PubMed=8635746;
RA Chen N., Upcroft J.A., Upcroft P.;
RT "A new cysteine-rich protein-encoding gene family in Giardia
RT duodenalis."
RL Gene 169:33-38 (1996).
DR EMBL; L39804; AAB06228.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 1.
DR SMART; SM00261; FU; 3.
DR PROSITE; PS01186; EGF_2; UNKNOWN_4.
SQ SEQUENCE 645 AA; 65262 MW; F19FE98DBB0AA589 CRC64;
```

Query Match 9.1%; Score 232; DB 5; Length 645;
Best Local Similarity 22.9%; Pred. No. 4.9e-09;
Matches 117; Conservative 40; Mismatches 183; Indels 170; Gaps 28;

```
QY 19 KSANCPVGTETNTAGVDLLCTPANCYNCKNFYNNAAAFVPGASTCTPC--POKKDAG 76
DB 176 KASNTDCGEELKRAG-----CATC-----TAVGPNQGTCLTTCNGGQVQLN 216
```

```

QY 77 AOPNPATANLVTCQNVKCPAGTATAGGATDYAAIITEVCNCRINFYN-----ENAPNFN 131
Db 217 G-----ISCGDSCPSNSAAMPD-----ICEC-NEGPNLNSGKDCCEKASNTD 257
QY 132 -----AGASTCTACPNVRVGGALTAGNAATIV-----AQCNAVCPGTALDDGGVTTDYV 180
Db 258 CGEELKRAAGCATCTAVGPN--GOTCLTCNGGQKVLNGISCGDSCPSNSAAMPDI----- 310
QY 181 RSFTECVKRLNFYNGNG-----NTPFNPGRSQCTPCPAIKPAN--VAQAT 226
Db 311 -----CECEGPNLNSGKDCCEKASNTQNTP-----NCKTCDNPKTDNEVCTECN 356
QY 227 LGNDATITACQNVACPDGTISAGVNNVAQNTNCAPNFYNNAPNFNPGNSTCLPC 286
Db 357 DGDYLTPTNOQVPDCT--TISGYVDN-----DKCKACSPEC-----AECVGPANNQCCSSC 406
QY 287 PANK--DYGAETAGATLAKOCNACPDGTATAS-----GATNYVI---LQTE----- 331
Db 407 PAGKKLTYYDDSNPNNGTCDGACKVSA--DGTGCTCGAIGGTGTAICSKCKTSTQAPLNG 465
QY 332 -----CLNCAANFYF-DG-----NNFQAGSSRCKACPA 358
Db 466 DCAASSRATFCTKMGNGVCTOCEDNYFLKDGCGYQTDROPKGQVCSNAQGGNGKQCT-- 523
QY 359 NKVOGAVATAGTATLIAQCALECPAGTVLTDGTTSYKQAAECVCKAANFYTTKQTDW 418
Db 524 --ANGLAATDGNCAECHPTCA-----TCSAPSTASSCKTCAATGYKENGDDT 568
QY 419 VAGIDTCTSCNKKLTSGAEANLPESAKKNI 448
Db 569 TDG--PCMCKSEKI-SGCKQCVSSSGSVI 595

RESULT 11
Q24977
ID Q24977 PRELIMINARY; PRT: 1274 AA.
AC Q24977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine rich protein.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen N., Upcroft P., Upcroft J.;
RT "A Giardia duodenalis gene encoding a protein with multiple repeats of
a toxin homologue.";
RL Parasitology 111:0-0(1995).
DR EMBL; L29079; AAA74587.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 1.
DR SMART; SM00181; EGF; 27.
DR SMART; SM00261; FU; 16.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 1274 AA; 135720 MW; 321622872A971A32 CRC64;

```

```

Query Match          9.0%; Score 228; DB 5; Length 1274;
Best Local Similarity 23.1%; Pred. No. 2e-08;
Matches 119; Conservative 35; Mismatches 183; Indels 178; Gaps 30;

QY 21 ANCPV-GTETNTAGQVDDLGPANCVNCQKNFYNN-----NAAAFVPGASCTPC 69
Db 801 APCNVGCECTVEGNAQ-----QKTCRPGYINTDTKQCTKDPEAPCNVEGCETCV-- 852
QY 70 POKKDAGAQP-----NPPATANLVQ-----CNVY-CPAGTATAGGATDYAAIITEC 115
Db 853 ----EGNAQQCKTCRPGYINTDTKQCTKDPEAPCNVEGCE--TCVEGNA-----QQC 899

```

```

QY 116 VNCRINFENAPNFNAGASTCTACPNVRVGGALTAGNAATIVACQNVACPTGTALDDGV 175
Db 900 KTCRPGY-----TINTDTKQCTKDP-----EAPCNV-----EGC 928
QY 176 TTDYVRSFTECVKRLNFYNGNG-----NTPFNPGRSQCTPCPAIKPANVAQA 225
Db 929 ETCVEGNAQQCKTCRPGYINTDTKQCTKDPEAPCNTP-----NCKTCDNPKTDNEI-C 981
QY 226 TLGNDATITACQNVACPDGTISAGV-----NNWVAQNTNCAPNFYNNAPNFNPGN 280
Db 982 TKCNDGDLTPTNOQVPDCT-AISGYVDGTDKKCKACNPECAECV-----GPN 1029
QY 281 STCLPCPANK--DYGAETAGATLAKOCNIA-----CPD-----GTAISGATN--- 324
Db 1030 NOCTACPVGKMLQYTDINTPVNGTCDQCVSVSTNDGCAECGAQIGGTAYCSKCKNTQ 1089
QY 325 -----YVILQTECLNCAANFYF-DG-----NNFQAGSSRC 353
Db 1090 APLNGNCAASSRVAFCATITSGACTKCNFYFLKDGCGYQTDROPKGQVCSNAQGGNGKC 1149
QY 354 KACPANKVOQAVATAGTATLIAQCALECPAGTVLTDGTTSYKQAAECVCKAANFYTT 413
Db 1150 QTC-----ANGLAASDGNCA-ECHS-----TCAFCST-ADAADCKCTCATGYKE 1192
QY 414 KQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNI 448
Db 1193 NGDDTTAGL--CKKCKSEKI-SGCKQCVSSSGSVI 1224

RESULT 12
Q9XTK3
ID Q9XTK3 PRELIMINARY; PRT: 667 AA.
AC Q9XTK3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE VSP417-3/A-II.
GN VSP417-3/A-II.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRIS/83/HEPU/136;
RX MEDLINE=99053029; PubMed=9836309;
RA Ey P.L., Darby J.M., Mayrhofer G.;
RT "Comparison of ts417-like variant-specific surface protein (VSP)
genes in Giardia intestinalis and identification of a novel locus in
genetic Group II isolates.";
RL Parasitology 117:445-455(1998).
DR EMBL; AF033584; AAD03497.1; -.
DR HSP; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00261; FU; 3.
SQ SEQUENCE 667 AA; 69123 MW; 77C64CFF59441C0C CRC64;

```

```

Query Match          9.0%; Score 227.5; DB 5; Length 667;
Best Local Similarity 24.3%; Pred. No. 1.1e-08;
Matches 120; Conservative 39; Mismatches 176; Indels 159; Gaps 33;

QY 18 IKSANCPVCTETNTAGQVDDLGPANCVN---CQKNFYNNAAAFV-----PG 62
Db 171 IKTATC-TGCDSNKIVKTDTSGT--SCTEESACSGFFVSDQQAQSKDYKCIPIRIDPA 227
QY 63 ASTCTPCPKQKADAGQNPANLVTCN-----TECNCTDQHCACFVAEGTCQKSSGILDG--- 275
Db 228 KANCITACSDNK---KPNLEG-----TECNCTDQHCACFVAEGTCQKSSGILDG--- 275

```


Db	453	AVLYKNDGT-KGTCGAGCATGTGAGA-----CKTCG-----LIIDGTSY	491
Qy	283	CLPCPANKDY-----GAEATAGGAATLAKQCNIACPDG-----	315
Db	492	CSECAVETEXPQGGVCSSTTVRAAATCKAGSVAGMGN-SCTNGLFMRMNGGCVETTKFPG	550
Qy	316	-----TAIASGATNVILQTECLNCAANFYFDGNNFQAGSRCKACQANKV-----QGAV	365
Db	551	KSVCEEAASAGDT-----COKEAPGYHLNNDLVTCSPGKCTCTSNIVCTACMEGYV	602
Qy	366	ATAGGTATLIAQCALECPAGTVLTDTSTYKQAASECVKCAANFYTTKQTD-WVAGIDT	424
Db	603	KTSDSCAKAAGCA-TCTGSGTTACDTCSTGYKSGTTVCSC-----TESNSDKTITGVAN	656
Qy	425	CTSCNKKLTS	434
Db	657	CAXCAPPLNN	666

RESULT 15
Q9NGL3 PRELIMINARY; PRT; 1274 AA.
ID Q9NGL3 AC Q9NGL3; DT 01-OCT-2000 (TrEMBLrel. 15, Created) DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update) DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update) DE variant-specific surface protein VSP136b. GN VSP136b. OS Giardia lamblia (Giardia intestinalis). OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia. OX NCBI_TaxID=5741; RN [1] RP SEQUENCE FROM N.A. RC STRAIN-AD-1; RA Mansouri M., Ey P.L.; RT "Analysis of a vsp136 homolog in Giardia intestinalis."; RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases. EMBL; AF249878; AAF69839.1; . DR InterPro; IPR000345; CytC_heme_bind. DR InterPro; IPR000561; EGF-like. DR InterPro; IPR002174; Furin-like. DR InterPro; IPR005127; Giardia_VSP. DR Pfam; PF03302; VSP; 1. DR SMART; SM00181; EGF; 23. DR SMART; SM00261; FU; 14. DR PROSITE; PS00190; CVTOCHROME_C; UNKNOWN 1. SQ SEQUENCE 1274 AA; 135671 MW; 68644A814BD6AE65 CRC64;

Query Match	8.6%	Score 219;	DB 5;	Length 1274;
Best Local Similarity	22.9%	Pred. No. 8.9e-08;		
Matches 118;	Conservative 35;	Mismatches 184;	Indels 178;	Gaps 30;
Qy	21	ANCPV-GTETNTAGQVDDLGPANCVCQKNFYNN-----NAAAFVPGASTCTPC	69	
Db	801	APCNVEGCTCEVNAQ-----QCKTCRPGYTTINTDKQCTKDPEAPCNVEGCTCV--	852	
Qy	70	POKDDAGAP-----NPPATANLVQ-----CNVK-CPAGTAIAGGATDYAAIITEC	115	
Db	853	-----EGNAQCKTCRPGYTTINTDKQCTKDPEAPCNVEGCE--TCVEGNA-----QQC	899	
Qy	116	VNCRINFYNENAPNAGASTCTACPNVRVGGALTAGNAATIVAOQNVACPTGTALDDGV	175	
Db	900	KTCRPGY-----TINTDKQCTKDP-----EAPCNV-----EGC	928	
Qy	176	TDYVRSFTECKRLNFYNNNG-----NTPFNGKSQCTPCPAIKPANVAQA	225	
Db	929	ETCVEGNAQCKTCRPGYTTINTDKQCTKDPEAPCNTP-----NCKTCDNPKTDNEI-C	981	
Qy	226	TIGNDATITAOQNVACPDGTISAAGV-----NNWVAQNTCTNCAPNFYNNAPNFNGN	280	
Db	982	TKCNDGDYLTPTNQCPDCT-AISGYIGDUDKCKACNPECAECV-----GPN	1029	